

## V2X 네트워크를 위한 DRL 기반 자원 블록 및 전송 파워 통합 제어

최희주, 채승호\*

한국공학대학교

{babybird, shchae}@tukorea.ac.kr

## DRL-based Integrated Resource Block and Power Control for V2X Networks

Heeju Choi, Seong Ho Chae\*

Tech University of Korea

## 요 약

안전이 특히 중요시되는 V2X(Vehicle-to-Everything) 환경에서는 V2V(Vehicle-to-Vehicle) 링크에 별도의 전송 자원을 할당해주는 것이 유용한 방안이 될 수 있다. 따라서 제한된 자원을 V2I(Vehicle-to-Infrastructure) 링크와 V2V 링크 사이에서 효율적으로 분할하고, V2V 링크들에 대한 RB(Resource Block) 할당 및 전력 제어 방안이 필요하다. 본 논문에서는 V2I 링크의 합산 데이터율을 최대화하면서 V2V 링크의 성공적 전송을 보장하는 DRL(Deep Reinforcement Learning) 기반 RB 분할·할당 및 전력 제어 알고리즘을 제안한다.

## I. 서 론

최근 자율주행 차량과 지능형 교통 시스템(Intelligent Transportation System, ITS)의 발전으로 차량과 주변 환경과의 통신을 지원하는 V2X(Vehicle-to-Everything) 기술이 주목받고 있다[1]. 다양한 V2X 통신 모드 중에서도 차량-인프라 간 통신(Vehicle-to-Infrastructure, V2I)과 차량 간 통신(Vehicle-to-Vehicle, V2V)은 핵심적인 역할을 담당한다. V2I 링크는 차량과 인프라 간의 고속 데이터 교환을 가능하게 하여, 비디오 스트리밍과 지도 업데이트와 같은 엔터테인먼트 애플리케이션을 지원한다. 반면, V2V 링크는 충돌 방지, 협력 주행, 긴급 알림과 같은 안전 관련 메시지를 전송하며, 이를 위해 초저지연과 높은 신뢰성이 요구된다[2]. 그러나, V2X 네트워크에서 한정된 자원으로 인한 효율적 자원 할당은 항상 중요한 문제이며, 이에 대한 연구들이 활발히 진행되었다. 대부분의 연구는 V2I 링크와 V2V 링크가 동일한 무선 자원을 공유하는 시나리오에 초점을 맞추고 있다. 반면 안전이 특히 중요시되는 환경에서는 V2V 링크에 별도의 전송 자원을 할당하여 안전 메시지의 성공적 전송을 보장하는 방안이 유용할 수 있다. 이는 제한된 무선 자원을 V2I 링크와 V2V 링크 사이에서 얼마나 효율적으로 분할·할당한 것인가라는 중요한 과제를 제기한다.

심층 강화 학습(Deep Reinforcement Learning, DRL)은 복잡한 환경에서도 학습과 적응이 가능하며, 환경 변화에 따라 자율적으로 최적의 자원 할당 전략을 탐색할 수 있는 강력한 도구로 주목받고 있다[3]. 이러한 장점 덕분에 V2X 네트워크에서도 자원 할당 문제를 해결하기 위한 효과적인 접근법으로 활발하게 연구되고 있다. 따라서 본 논문에서는 DRL을 활용하여 V2X 네트워크에서 V2I 링크와 V2V 링크 간 자원 블록(Resource Block, RB) 분할 비율을 실시간으로 최적화하고, V2V 링크의 RB 할당과 전력 제어를 효과적으로 수행하는 새로운 알고리즘을 제안한다.

## II. 시스템 모델

우리는 교차로 중앙에 기지국(Base Station, BS)이 위치하고, V2I 링크를 형성하는  $M$ 개의 V2I-VUE와 V2V 링크를 형성하는  $N$ 개의 V2V-VUE 쌍(pair)이 분포해 있는 V2X 네트워크를 고려한다. 네트워크는 이산 시간 슬롯  $t \in \{1, 2, \dots, T\}$  단위로 동작하며, 차량들은 매 슬롯마다 차선을 따라 위치를 바꾼다. 전체  $K$ 개의 RB를 V2I 링크 전용  $K_1(t) = (1 - \lambda(t))K$ 개와 V2V 링크 전용  $K_2(t) = \lambda(t)K$ 개로

분할하여, V2I 링크와 V2V 링크 간에 간섭이 없도록 한다. 여기서  $\lambda(t) \in \{0, 1/K, \dots, 1\}$ 는 RB 분할 비율이다. V2I 링크의 경우, 평균 수신 전력이 높은 상위  $K_1(t)$ 개의 V2I-VUE에만 RB를 할당하고, V2V 링크의 경우, 각 링크가  $K_2(t)$ 개의 RB 중 하나를 할당받는다.

V2I-VUE  $m$ 의 RB  $k$ 에서의 SNR (Signal-to-Noise Ratio)과 V2V-VUE 쌍  $n$ 의 RB  $k$ 에서의 SINR (Signal-to-Noise-plus-Interference Ratio)은 각각 아래와 같이 주어진다.

$$\gamma_{m,k}^I(t) = \frac{\delta_{m,k}(t)P_m^I h_{m,k}(t)}{\sigma^2}, \quad (1)$$

$$\gamma_{n,k}^V(t) = \frac{\rho_{n,k}(t)P_n^V g_{n,k}(t)}{\sigma^2 + \sum_{j \in \mathcal{N} \setminus \{n\}} \rho_{j,k}(t)P_j^V g_{j,n,k}(t)}, \quad (2)$$

여기서,  $\delta_{m,k}(t) \in \{0, 1\}$ 는 시간  $t$ 에서 RB  $k$ 가 V2I-VUE  $m$ 에 할당되었는지를,  $\rho_{n,k}(t) \in \{0, 1\}$ 는 시간  $t$ 에서 RB  $k$ 가 V2V-VUE 쌍  $n$ 에 할당되었는지를 나타내는 이진 지시변수이다.  $P_m^I$ 은 V2I-VUE  $m$ 의 고정 송신 전력이고,  $P_n^V(t)$ 는 시간  $t$ 에 V2V-VUE 쌍  $n$ 이 사용하는 송신 전력이다.  $h_{m,k}(t)$ 와  $g_{n,k}(t)$ 는 각각 V2I-VUE  $m$ 과 V2V-VUE 쌍  $n$ 의 채널 전력 이득이며,  $g_{j,n,k}(t)$ 는 V2V-VUE 쌍  $j$ 의 송신기가 V2V-VUE 쌍  $n$ 의 수신기에 미치는 간섭 채널 전력 이득을 의미한다. 마지막으로,  $\sigma^2$ 은 채널 잡음 전력이다.

따라서, V2I-VUE  $m$ 의 RB  $k$ 에서의 성취가능한 데이터율과 V2V-VUE 쌍  $n$ 의 RB  $k$ 에서의 성취가능한 데이터율은 각각 아래와 같이 주어진다.

$$C_{m,k}^I(t) = W \log_2(1 + \gamma_{m,k}^I(t)), \quad (3)$$

$$C_{n,k}^V(t) = W \log_2(1 + \gamma_{n,k}^V(t)), \quad (4)$$

여기서,  $W$ 는 각 RB의 대역폭이다.

모든 V2V 링크가 주어진 시간 내에 페이로드를 전송해야 하므로, 본 논문의 목표는 V2I 링크의 성취가능한 합산 데이터율을 최대화하면서 V2V 링크의 전송 성공을 보장하는 최적의 전체 RB 분할값,  $\lambda(t)$ , 각 V2V 링크에 대한 RB 할당,  $\rho_{n,k}(t)$ , 각 V2V 링크에 대한 전력 제어,  $P_n^V(t)$ 을 정하는 정책을 찾는 것이다. 최적화 문제는 다음과 같이 쓸 수 있다.

$$P: \max_{\{\lambda(t)\}, \{\rho_{n,k}(t)\}, \{P_n^V(t)\}} \sum_{t \in T} \sum_{m \in M} \sum_{k \in K_1(t)} C_{m,k}^I(t), \quad (5)$$

$$\text{s.t.} \sum_{t \in T} \sum_{k \in K_2(t)} C_{n,k}^V(t) \cdot \Delta t \geq L, \quad \forall n \in N, \quad (6)$$

$$\sum_{k \in K_2(t)} \rho_{n,k}(t) \leq 1, \quad \forall n \in N, \quad \forall t \in T, \quad (7)$$

$$\rho_{n,k}(t) \in \{0,1\}, \quad \forall k \in K_2(t), \quad \forall t \in T, \quad (8)$$

여기서,  $\Delta t$ 는 한 시간 슬롯의 지속 시간을 나타내며,  $L$ 은 주기적으로 생성되는 V2V 링크 페이로드의 크기이다. 제약 조건 (6)은 각 V2V 링크가 주어진 시간 내에 모든 페이로드를 전송하도록 보장하고, 제약 조건 (7)은 각 V2V 링크가 최대 하나의 RB만 사용하도록 한다. 마지막으로 제약 조건 (8)은 모든 V2V 링크가 V2V 링크 전용 RB 중 하나를 할당받도록 보장한다.

### III. 제안하는 DRL 기반 RB 분할, 할당, 전력 제어 알고리즘

본 장에서는 V2X 네트워크에서 DRL을 활용하여 V2V 링크의 안전 메시지의 성공적 전송을 보장하면서 V2I 링크의 합산 데이터율을 최대화하는 방법을 제안한다. RB 분할, 할당 및 전력 제어 문제는 다음의 마르코프 결정 과정(Markov Decision Process, MDP)으로 정의된다.

- 1) 에이전트(Agent): BS가 에이전트로 학습을 진행한다.
- 2) 상태(State): 에이전트는 현재 모든 V2V 링크의 남은 페이로드 크기, 전송 완료까지 남은 시간, 모든 V2I 링크와 V2V 링크의 채널 이득, 모든 V2V-VUE의 송·수신기 간 거리, 이전 시간 슬롯에서의 자원 분할 비율 값을 상태로 인식한다.
- 3) 행동(Action): 에이전트는 전체  $K$ 개의 RB를 어떻게 V2I 링크용과 V2V 링크용으로 분할할지 정하고, 각 V2V 링크에 대해 어떤 RB를 할당할지와 송신 전력 값을 결정한다.
- 4) 보상(Reward): 보상은 크게 두 부분으로 구성된다. 첫 번째 부분은 V2I 링크의 전체 데이터율에 비례한 보상으로, 에이전트가 자원 할당을 통해 얼마나 높은 V2I 링크 합산 데이터율에 달성했는지를 그대로 반영한다. 두 번째 부분은 V2V 링크의 전송 성공 여부에 따른 보상으로, 각 V2V 링크의 순간 전송 속도가 '남은 데이터량을 남은 시간 내에 전송할 수 있는 속도'를 넘으면 고정된 양의 보상값을 주고, 그렇지 못하면 음의 보상값을 부과한다. 이 두 보상 항목에 적절한 가중치를 부여해 합산함으로써, 에이전트가 V2I 링크의 성능을 높이면서도 V2V 링크의 전송 성공을 달성하도록 설계한다.

RB 분할, 할당 및 전력 제어 문제는 위의 MDP를 기반으로 DQN(Deep Q-Network) 알고리즘[5]을 활용하여 해결 가능하다. 페이지 제약으로 구체적인 알고리즘은 생략한다.

### IV. 성능 검증

본 시뮬레이션 환경에서는 3GPP TR 36.885[5]에 정의된 도시 도로 환경을 사용하며, DQN은 각각 512개, 256개, 128개의 뉴런을 가진 3개의 완전 연결 은닉층으로 구성되어 있다. 그림 1은 DQN 학습 과정에서의 각 에피소드마다 누적 보상을 나타낸다. 이를 통해 학습을 진행함에 따라 누적 보상값이 증가하는 경향을 확인할 수 있다.

제안된 알고리즘의 성능을 고정 RB 분할 비율(V2I 링크 전용:V2V 링크 전용 = 0:5, 1:4, 2:3, 3:2, 4:1, 5:0)을 가진 랜덤(Random) 기반 RB 할당 및 전력 제어 알고리즘과 비교하였다. 표 1은 V2V 링크의 안전 메시지 크기가  $9 \times 1060$  Bytes인 경우의 성능 비교표를 나타내며, 제안된 알고리즘이 비교 알고리즘 대비 V2V 링크 전송 성공을 보장하면서 동시에 높은 V2I 링크의 총 데이터율을 달성하는 것을 확인할 수 있다.

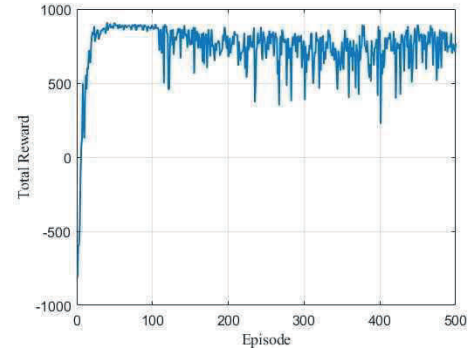


그림 1. 학습 과정에서의 누적 보상

Scheme	V2V payload transmission success probability	Achievable rate of V2I-VUEs(Mbps)
0:5 & Random	0.7000	0.000
1:4 & Random	0.5375	2.263
2:3 & Random	0.4125	4.445
3:2 & Random	0.2625	6.538
4:1 & Random	0.0000	8.261
5:0 & Random	0.0000	9.674
Proposed	1.0000	4.483

표 1. 성능 비교표

### V. 결 론

본 논문에서는 V2X 네트워크에서 V2I 및 V2V 링크를 위한 DRL 기반 RB 분할, 할당 및 전력 제어 알고리즘을 제안하였다. 제안된 알고리즘은 전체 RB를 V2I 링크 전용 RB와 V2V 링크 전용 RB로 동적으로 분할하고, V2V 링크 전용 RB를 V2V 링크에게 효율적으로 할당하고 전력 제어를 함으로써 V2I 링크의 합산 데이터율을 최대화하면서도 V2V 링크 통신의 신뢰성을 보장한다. 시뮬레이션 결과, 제안된 DRL 기반 알고리즘이 비교 알고리즘 대비 V2V 링크 전송 성공을 보장하면서 V2I 링크의 총 데이터율 측면에서도 우수함을 보였다.

### ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-학·석사 연계ICT핵심인재양성의 지원(IITP-2025-RS-2022-00156326, 50%)과 정보통신기획평가원-지역지능화혁신인재양성사업의 지원을 받아 수행된 연구임(IITP-2025-RS-2020-II201741, 50%)

### 참 고 문 헌

- [1] H. Ye, G. Y. Li and B. -H. F. Juang, "Deep reinforcement learning based resource allocation for V2V communications," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 3163-3173, Apr. 2019.
- [2] X. Zhang *et al.*, "Deep-reinforcement-learning-based mode selection and resource allocation for cellular V2X communications," *IEEE Internet Things J.*, vol. 7, no. 7, pp. 6380-6391, Jul. 2020.
- [3] P. Xiang *et al.*, "Multi-agent reinforcement learning-based decentralized spectrum access in vehicular networks with emergent communication," *IEEE Commun. Lett.*, vol. 27, no. 1, pp. 195-199, Jan. 2023.
- [4] C. Lee *et al.*, "Multi-agent deep reinforcement learning based handover strategy for LEO satellite networks," *IEEE Commun. Lett.*, vol. 29, no. 5, pp. 1117-1121, May 2025.
- [5] *Technical specification group radio access network: study LTE-based V2X services: (Release 14)*, document 3GPP TR 36.885 V14.0.0, 3rd Generation Partnership Project, Jun. 2016.